

METHOD AND SYSTEM FOR IDENTIFYING A USER BY VOICE

Attorney Docket Number

1094

Express Mail Label Number

5 EL023800503US

Inventors

Sue McNeill

Robert C. Wohlsen

Field of the Invention

10 The present invention is related to computer software and more specifically to computer software for voice recognition.

Background of the Invention

In order to identify a user to a computer system,
15 voice recognition software may be used. A user may speak one or more forms of identification to the voice recognition software, and the voice recognition software attempts to identify the user using what was said to the voice recognition software. The system also attempts to
20 validate the identification using unique characteristics of the user's voice that have previously been identified during a "enrollment" procedure. The enrollment procedure enrolls the unique characteristics of the person's voice at

a time when the person has been positively identified using means other than voice, such as visual identification or a PIN number.

One approach described in U.S. Patent Number 5,517,558
5 issued to Schalk is illustrative. A user speaks, one character at a time, a password in response to a prompt. The password must be spoken one character at a time because single character recognition can be more accurate than whole word recognition for reasons such as the limited
10 number of single characters. The system uses speaker-independent voice recognition techniques to recognize the characters spoken as a password. During this identification process, the system also extracts parameters from the password spoken that identify the unique details
15 of the user's voice. These parameters extracted during the identification process are matched against parameters previously extracted during an enrollment process for the user having the password recognized.

If the two sets of parameters match relatively
20 closely, there is a high probability that the user is the user identified. If the two sets of parameters have little in common, the user may be rejected, and if the two parameters are in between a close match and little in

common, Schalk prompts for additional information, also to
be spoken one character at a time. Speaker independent
techniques are used to recognize the additional
information, and a match is attempted against additional
5 information stored for the user to validate the identity of
the user. However, in the preferred embodiment, Schalk
always prompts the user for the additional string and uses
speaker independent voice recognition techniques to confirm
the identity of the individual by matching the second
10 string to corresponding information stored for the user.

There are several problems with this approach. First,
each password in the system must be unique because the
password is used to uniquely locate the characteristics of
the person's voice. To enforce the rule that all passwords
15 are unique, passwords must be assigned rather than chosen,
or some users will not get their first password choice and
must select another password. In such systems, users who
cannot use their first choice as a password tend to forget
their passwords more often than a system in which every
20 user can use their own first choice password, even if such
password is already being used by another user. Second,
the Schalk system requires passwords to be spoken one
character at a time. Users may find such a system
unnatural to use.

What is needed is a method and apparatus that can identify a user without requiring that all passwords be unique or spoken one character at a time.

Summary of Invention

5 A method and system allows users to enroll by speaking their name and password. A grammar is extracted from the name and password that identifies how the user speaks the name and password. A voiceprint is extracted from the name and password that identifies characteristics of the user's
10 voice. The user's account can be marked as having been enrolled, and the voiceprint and grammar are stored associated with the user's account identifier. When a caller wishes to identify himself as a user, he is prompted for his name and password, and speaks his name and
15 password. The method and system uses conventional voice recognition on the name to narrow the list of possible users from those who have enrolled to a smaller number of users most closely matching the recognized name. The method and system extracts the grammar from the caller's
20 password and attempts to match the grammar against the grammars associated with the smaller number of users to identify an even smaller number of users or as few as one user. The method and system extracts a voiceprint from the

name and grammar and matches the voiceprint extracted with the voiceprints of the smaller number of users obtained during the enrollment process. The match is performed either to narrow the identity of the caller from the smaller number of users to a single user, or to verify the identity of the single user whose grammar of the password matches that of the caller.

Brief Description of the Drawings

Figure 1 is a block schematic diagram of a conventional computer system.

Figure 2 is a block schematic diagram of a system for identifying a person to a computer according to one embodiment of the present invention.

Figure 3A is a flowchart illustrating a method of a method of enrolling a person to a computer according to one embodiment of the present invention.

Figure 3B is a flowchart illustrating a method of identifying a caller as a user according to one embodiment of the present invention.

Detailed Description of a Preferred Embodiment

The present invention may be implemented as computer software on a conventional computer system. Referring now

to Figure 1, a conventional computer system 150 for practicing the present invention is shown. Processor 160 retrieves and executes software instructions stored in storage 162 such as memory, which may be Random Access Memory (RAM) and may control other components to perform the present invention. Storage 162 may be used to store program instructions or data or both. Storage 164, such as a computer disk drive or other nonvolatile storage, may provide storage of data or program instructions. In one embodiment, storage 164 provides longer term storage of instructions and data, with storage 162 providing storage for data or instructions that may only be required for a shorter time than that of storage 164. Input device 166 such as a computer keyboard or mouse or both allows user input to the system 150. Output 168, such as a display or printer, allows the system to provide information such as instructions, data or other information to the user of the system 150. Storage input device 170 such as a conventional floppy disk drive or CD-ROM drive accepts via input 172 computer program products 174 such as a conventional floppy disk or CD-ROM or other nonvolatile storage media that may be used to transport computer instructions or data to the system 150. Computer program product 174 has encoded thereon computer readable program

code devices 176, such as magnetic charges in the case of a floppy disk or optical encodings in the case of a CD-ROM which are encoded as program instructions, data or both to configure the computer system 150 to operate as described
5 below.

In one embodiment, multiple computer systems 150 are used to implement the present invention. A conventional mainframe computer such as a conventional S/390 computer system commercially available from IBM Corporation of
10 Armonk, New York may be coupled to one or more conventional Sun Microsystems Ultra Sparc computer systems running the Solaris 2.5.1 operating system commercially available from Sun Microsystems of Mountain View, California, although other systems may be used. A VPS recognizer commercially
15 available from Periphonics Corporation of Bohemia, New York and any of Nuance 6, Nuance Verifier, Nuance Developer's ToolKit and Speech Objects software commercially available from Nuance Communications of Menlo Park, California are used with the Ultra Sparc computer to perform certain
20 recognition functions.

System - Enrollment

Referring now to Figure 2 a system 200 for identifying a person to a computer is shown according to one embodiment

of the present invention. In one embodiment, a user enrolls with the system 200 during one session, then identifies himself to the system 200 during the same or different session (between sessions, the user may be
5 disconnected from the system 200). To enroll with the system, a user is vocally connected with enrollment manager 220 at input 210. Input 210 may be reached by dialing it or via a studio or transfer from an operator who receives the caller and verifies the callers identity. If the
10 caller dialed in, enrollment manager 220 may validate the identity of the caller by prompting for one or more forms of identification, such as an account number and PIN. This information may be spoken or keyed in using touch tones or other input methods. Enrollment manager 220 stores this
15 information internally for use as described below.

To validate the identity of the caller, enrollment manager 220 can request from a account information storage 236 which may be a conventional personal computer, minicomputer or mainframe computer, the corresponding
20 identification for the user. For example, enrollment manager 220 can request from account information storage 236 the PIN corresponding to the account number received from the user. Enrollment manager 220 can verify the identity of the user by comparing the PIN received from

account information storage 236 with the PIN received from the user. If enrollment manager 220 verifies the identity of the user, enrollment manager 220 prompts the user to separately speak his name and password in one embodiment.

5 Although a name and password are used herein in one embodiment, other forms of identification may be used, and a larger or smaller number of forms of identification may also be used in other embodiments.

10 In one embodiment, enrollment manager 220 passes the name and password spoken by the user to voiceprint extractor 222 for use as described below, and passes the name and password to grammar extractor 224 for use as described below. In another embodiment, enrollment manager 220 passes only the password to grammar extractor 224.

15 Voiceprint extractor 222 receives the spoken name and password from enrollment manager 220 and can receive other spoken information as well. For example, in one embodiment, enrollment manager 220 prompts each user to speak the same phrase, such as "My voice is my password"

20 and provides the response to voiceprint extractor 222 to for extraction. Voiceprint extractor 222 uses conventional speaker verification modeling techniques to extract one or more characteristics and/or patterns of the user's voice

that can uniquely identify the user from the general population or at least discriminate a user from at least approximately 99.5% of the general population with near zero false acceptances and near zero false rejections.

5 Voiceprint extractor 222 stores in voiceprint storage
232 the voiceprint, extracted as described above. In one
embodiment, voiceprint extractor 222 receives from
enrollment manager 220 the account number or other
10 identifier of the user. Voiceprint extractor 222 stores
the voiceprint associated with the account number or other
15 identifier of the user whose voice the voiceprint
describes. This may be performed using a database or by
storing each voiceprint in a separate file with a filename
equal to the account number or using any other conventional
technique.

Grammar extractor 224 receives the user's spoken
password from enrollment manager 220, and extracts a
grammar from the password. Grammars are a description of
the sounds and transitions between them and order of the
20 sounds that form a word or words. In one embodiment, a
grammar is similar to a dictionary pronunciation of a word,
which describes the elemental sounds that make up the word,
and provides an indication of emphasis, implying loudness

and pitch relative to the rest of the word. If grammar extractor 224 receives the name from enrollment manager 220, grammar extractor 224 extracts a grammar from the name.

5 Grammar extractor 224 stores the grammar of the user's spoken password in grammar storage 234. In one embodiment, grammar extractor 224 receives the user's account number or other identifier from enrollment manager 220, and associates the grammar or grammars it stores in grammar
10 storage 234 with the account number or other identifier as described above with reference to voiceprint extractor 222. This may be accomplished using conventional database techniques or other conventional techniques.

If grammar extractor 224 receives the user's name,
15 grammar extractor 224 repeats the process above on the name, storing the grammar of the name in grammar storage 234.

In one embodiment, when voiceprint extractor 222 and grammar extractor 224 complete the process described above,
20 the signal enrollment manager 220. Enrollment manager 220 marks in account information storage 236 the name of any user that successfully completes the enrollment procedure described above so that the recognition process described

below need only attempt to recognize those names of people who have enrolled.

If desired, the prompting process described above for the name, password or both may be repeated a number of
5 times such as three, and the responses received used to refine the extraction of the voiceprint, grammar or both. Other information may also be prompted for and received to refine the extraction described above.

System - Identification

10 When a user wishes to be identified using his voice, the user connects to ID manager 240 via input/output 212. Input/output 212 may be a conventional dial-up telephone line. ID manager 240 contains a conventional telephone interface, although other means of interface such as an
15 Internet telephony interface may be used.

When ID manager 240 receives a call, ID Manager 240 signals error/name prompter 242 which prompts the caller to speak his or her name via input/output 212. ID manager 240 listens for a response and passes voice responses to name
20 recognizer 244 and to voiceprint extractor 222. ID manager 240 may pass any responses in analog or digital form and may pass the response in real time or after recording them.

Before the name is spoken, the caller's account number could be any of the valid accounts corresponding to those people who have enrolled. Name recognizer 244 uses conventional voice recognition techniques attempts to

5 narrow the possible accounts corresponding to the name spoken down to a smaller number of accounts than all of the accounts of people who have enrolled, but greater than one account. For example, name recognizer 244 can attempt to narrow the list down to ten accounts or even a hundred,

10 which may be smaller than the number of valid, enrolled accounts, which may number in the thousands, tens of thousands, hundreds of thousands or millions.

Name recognizer 244 can use any of speaker dependent or speaker independent techniques to try to identify the

15 several accounts that potentially correspond to the name it receives. For example, it can send the name to grammar extractor 224 to extract the grammar of the name received and then try to find in grammar storage 234 the nearest matches of grammars of the people who have spoken their

20 names. Alternatively, name recognizer 244 can attempt to use speaker independent voice recognition techniques to narrow the list of possible accounts down to a smaller number. Each of these techniques will now be described in more detail.

In one embodiment, name recognizer 244 sends to grammar extractor 224 the name it receives from ID manager 240. Grammar extractor 224 extracts the grammar from the spoken name received from name recognizer 244 and passes it
5 back to name recognizer 244. Name recognizer 244 attempts to locate in grammar storage 234 using conventional pattern matching techniques the N best matches between the grammar it receives from grammar extractor 224 and the grammars for the names that were extracted during the enrollment
10 procedure for each user as described above. N may be 10 or 100 or any other number.

In an alternate embodiment name recognizer 244 uses the text of the names of all enrolled users that are stored in account information storage 236 to perform the
15 recognition using conventional speaker independent speech recognition techniques. Name recognizer 244 attempts to resolve the name into a sequence of primitive sounds, then identifies possible spellings of those sounds. In one embodiment, name recognizer 244 locates from among the
20 names marked as enrolled the N best matches with the name spoken by the caller in response to the prompt by comparing the possible spellings with the actual spellings of the names of the enrolled users stored in account information storage 236. A similar technique may be performed by

having name recognizer 244 identify and store the primitive sounds of the name of each enrolled account stored in account information storage and store this information in account information storage 236 for comparison with the primitive sounds received. Other speaker independent techniques may also be used.

Speaker independent recognition may make use of utterance information spoken by a wide variety of people using a wide variety of accents and voices via input 245.

10 Name recognizer 244 uses this information to identify each primitive sound using conventional techniques.

Another speaker independent technique is to have name recognizer 244 compare the name received to a vocabulary of known names. The feasibility of this technique will depend on the number of possible names. As noted above, the names may be replaced with other words for which a vocabulary may be more feasible.

15

In one embodiment, N may be 10, in another embodiment, N may be 100, and N may be other numbers in still other embodiments. In an alternative embodiment, instead of requiring N matches to occur, a confidence level is selected, and name recognizer 244 locates all matches exceeding the confidence level. In any of the embodiments,

20

the names corresponding to more than one account are identified as a potential match for the name that is spoken in response to the prompt. In one embodiment, name recognizer 244 indicates the account names that are
5 potential matches by marking in account information storage 236 the account records corresponding to the names it identifies.

Name recognizer 244 signals ID manager 240 when it has performed the functions identified above. ID manager
10 240 signals password prompter 250, which prompts the user for his or her password via input/output 212. ID manager 240 passes any responses to grammar extractor 224 and to voiceprint extractor 222, and signals speaker dependent recognizer 254.

15 Grammar extractor 224 extracts the grammar from the response to the password prompt and stores it in a special location in grammar storage 234. Speaker dependent recognizer 254 compares the grammar in the special location of grammar storage 234 with each of the grammars
20 corresponding to the accounts marked in account information storage 236 by name recognizer 244 as described above. Speaker dependent recognizer 254 identifies the closest match between the grammar in the special location and the

grammars corresponding to the marked accounts and passes an identifier of the account to ID manager 240. Because grammar extractor 224 extracts the grammar of the word, which describes how the word is spoken, it does not need to
5 recognize the word. Thus, the name and/or password need not be a real word.

As shown and described herein, grammar extractor 224 is used for both the enrollment process and the identification process. However, different grammar
10 extractors could be used for each process.

Because ID manager 240 passes the name and password to voiceprint extractor 222, simultaneously with the operation of the name recognizer 244 and speaker dependent recognizer 254 and the ancillary activities (prompting, grammar
15 extracting, etc.) described herein, voiceprint extractor 222 extracts a voiceprint for the caller as described above, and stores the voiceprint so extracted in a special location of voiceprint storage 232.

ID manager 240 passes the identifier it receives from
20 speaker dependent recognizer 254 to verifier 252. Verifier 252 receives from ID manager 240 the identifier of the account ID manager 240 received from speaker dependent recognizer 254, and verifies the identity of the caller by

attempting to match some or all of the voiceprint extracted from the caller at input 212 with the voiceprint stored in voiceprint storage 232 for the account having the identifier verifier 252 receives from ID manager 240. If
5 the portion or all of the voiceprints match or nearly match within an acceptable tolerance level, verifier 252 provides the identifier of the account at output 260 to identify the caller. If the voiceprints do not match, verifier 252 signals ID manager 240. The acceptable tolerance level can
10 vary based on the application and the amount information collected from the person's voice, and can be set based on experience. In one embodiment, the confidence level will range from +1, indicating a highly probable match to -1 indicating no match. The acceptable confidence level is
15 +0.7 in one embodiment, although other values such as 0.6, 0.5, 0.8 and 0.9 may be used.

As shown and described herein, the same voiceprint extractor 222 is used for both the enrollment of the user and the verification of the user. However, different
20 voiceprint extractors (not shown) may be used for each of these functions.

In the event of an error, ID manager signals error/name prompter 242. An error occurs if name

recognizer 244 does not match any names (and therefore signals ID manager 240), speaker dependent recognizer 254 does not match any grammars (and therefore signals ID manager 240), or verifier 252 does not match the

5 voiceprints (and therefore signals ID manager 240), each as described above. Error/name prompter 242 informs the caller of the error and restarts the procedure above by prompting for the name.

Referring now to Figure 3A, a method of enrolling a
10 person to a computer is shown according to one embodiment of the present invention. A user is prompted for a name and an utterance that is a name is received 310 as described above. A voiceprint is extracted from the name and optionally a grammar or other description of the spoken
15 attributes of the name may be extracted 312 as described above.

The user is prompted for his or her password and an utterance containing the password is received 314 as described above. The voiceprint and grammar are extracted
20 from the password 316, 318 as described above. The information extracted in steps 312, 316, and 318 is stored, associated with the account number or other identifier of the user 320. The user's account may be marked 320 as

having successfully completed enrollment so that recognition may be attempted only on those accounts that are so marked as described in more detail above and below.

The voiceprint described in steps 312 and 316 may be
5 extracted and refined in these steps, and the caller may be prompted for the same or other information, which may be used to further refine the voiceprint. For example, the caller may be prompted for his or her name repeatedly in steps 310 and receipt of multiple copies of the name may be
10 used in the extraction of step 312. The password may also be repeatedly prompted in step 314 and each response used in steps 316, 318.

Referring now to Figure 3B, a method of identifying a caller as a user is shown according to one embodiment of the present invention. The caller is prompted for his or
15 her name and an utterance that is supposed to be the name is received 350 in response as described above. In one embodiment, step 350 includes distinguishing the name from background noise, coughs, static or other sounds that are
20 not words by using conventional speech recognition techniques. A voiceprint is extracted 352 from the name received in step 350. In addition, the name received in step 350 is recognized 354 using speaker independent or

speaker dependent techniques as described above to identify
356 the best N matches, where N is greater than one. N may
be ten or 100 or any other number that is smaller than
either the number of accounts or the number of accounts
5 that were marked as having completed enrollment as
described above. In one embodiment, the recognition is
attempted only for those accounts that were so marked, and
other accounts are skipped during the recognition step 354.
Step 356 may comprise marking the accounts of the
10 identified users.

If no accounts are identified or none are identified
where the recognition exceeds a specified confidence level,
an error occurs 358, the error may be explained 372 to the
user via an error prompt and the method continues at step
15 372. If desired, an error counter may be set to zero in a
step (not shown) preceding step 350 and incremented at step
372. When the error count reaches a specified threshold,
the caller is transferred to an operator or dropped. If no
error occurs 358, the method continues at step 360.

20 The caller is prompted for a password, and the
password is received 360 as described above. Additional
voiceprint information is extracted 362 from the password
received in step 360 and this additional voiceprint

information is added to, or used to refine, the voiceprint extracted in step 352 as described above.

The grammar is extracted 364 from the password received in step 360 as described above. The grammar
5 extracted in step 364 is compared with the grammars extracted as described above with reference to Figure 3A and the best match 366 with a single account from among those identified in step 356. Step 366 may be performed by
10 picking the closest match, above a specified confidence level (e.g. 95% match) to the grammar extracted in step 364.

If such a match is not located 368, the method continues at step 372 to allow a suitable error message to be played to the caller and the method may continue at step
15 350 in one embodiment, or at step 360 in another embodiment not shown. If the method continues at step 350, the accounts identified in step 356 have the identification removed (e.g. by removing the mark set in step 356 described above). If a match is located, 368, a
20 verification occurs as shown in steps 370-374 described below.

The voiceprint of the account identified as a match in step 366 is compared 370 with the voiceprint produced in

steps 352 and 362. The voiceprint of the account was produced as described above with reference to Figure 3A. If the voiceprints match 372, the account ID is provided 374 or another indication of authorization is provided in 5 step 374. Otherwise, the method continues at step 372 in the same manner as described above, with an indication of the error, and continuing at step 350 or 360.

In one embodiment not shown, if the "no" branch of step 372 is taken, an error counter flag initialized to a 10 value of 0 in a step (not shown) prior to step 350 is incremented at step 372. If the counter is equal to 1, the method continues at step 360. If the counter is equal to 2, the method continues at step 372. If the counter is equal to 3, the method continues by transferring the caller 15 to an operator, and optionally, the account matched in step 366 may be locked from further use.

In an alternate embodiment of the present invention, instead of narrowing the list of users down to a single user when the grammars are compared in step 366, and then 20 verifying the identity of the user by comparing the voiceprints, the list of users may be a small number, e.g. 5 after the grammars are compared in step 366, and the

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100	